

# Proof analysis using fine-grained dependency information in formal mathematics

**Jesse Alama**

Center for Artificial Intelligence  
New University of Lisbon

*Mathematical Logic in  
the Netherlands 2011*

2011-05-20

# Introduction

- Interest: representing mathematical knowledge formally and computing with it.
- Aim: discover necessary and sufficient conditions for (formalized) mathematical theorems
- Method: produce fine-grained dependency information from the **mizar** Mathematical Library (MML)



# Focus on mizar

- **mizar** is one of the oldest ITP that enjoys widespread use.
- Its foundations—classical first-order set theory—are the most attractive from the perspective of traditional foundations of mathematics.
  - **mizar** is actually based on Tarski-Grothendieck set theory, which is stronger than ZFC.
- Rich formalism: captures many natural aspects of mathematical practice, such as various kinds of notational conventions and extensions by definitions.
- Moreover, among comparable ITPs, the lion's share of the knowledge formalized in the mizar system is based on *pure mathematics*.



# Aim: computing precise necessary and sufficient conditions for (formalized) theorems

- From a formalized proof of a theorem, we can say, with complete precision, what is *sufficient* for the theorem.
- We can also compute refinements to a theorem, thereby giving a sharper set of sufficient conditions.
- When the set  $S$  of sufficient conditions is suitably refined, we can attempt to go backwards and find out whether  $S$  is actually necessary for the theorem ('reverse engineering mathematics')



# Dependency graph of the MML

- *Nodes*: **mizar** items (theorems, definitions, lemmas, implicit type information, etc.)
- *Edges*: The set of outgoing edges of a node  $u$  is the set of nodes  $\{v_1, v_2, \dots, v_n\}$  such that the justification of  $u$  depends on all of  $v_1, v_2, \dots, v_n$  to be successful according to the **mizar** verifier.

The dependency graph has 100,702 nodes and 4,175,861 edges. (And as the MML grows, these numbers will increase.)



# Philosophical aside: intensional vs. extensional mathematical knowledge

- Intension: the method of dependency graphs can be used to say something about new proofs *in mizar*:
  - Assumptions used in a proof of a mathematical theorem *as formalized in mizar*;
  - Necessary conditions for the success of a particular **mizar** text.
- Extension: claims about mathematical theorems that do not depend on how they are expressed/formalized.
  - Evidently something more is needed to answer extensional questions than working with a **mizar** dependency graph.
  - Candidates methods: 'thinking hard'/traditional mathematics, automated theorem provers.



# Proof refinement using dependency graphs

(example)

- Every path commencing with theorem  $T$  ending at axiom  $A$  passes through lemma  $L$ .
- Thus,  $S - \{A\} + L \vdash T$ .
- If we know that  $S - \{A\} + L$  differs from  $S$ , by our calculation we have found a weakening of the sufficient conditions for theorem  $T$ .



# Foundations of mizar: Tarski-Grothendieck set theory

- The MML is based on Tarski-Grothendieck set theory (**TG**), a classical first-order theory of sets in the signature  $\{=, \in\}$  whose axioms are:
  - extensionality
  - pairing
  - union
  - foundation/regularity
  - *Tarski's universe axiom*
- **TG** is equivalent to **ZFC** plus the principle that there are arbitrarily large strongly inaccessible cardinals.



# Tarski's universe axiom

- For every set  $a$  there exists a set  $U$  such that:
  - $a \in U$ ,
  - $U$  is closed under taking power sets,
  - $U$  is closed under union,
  - for every subset  $A$  of  $U$  for which  $|A| < U$ , we have that  $A \in U$
- Consequences of the universe axiom:
  - axiom of infinity
  - axiom of choice
  - power set



# Application of fine-grained dependency information: a discussion about AC and foundation in TG

- A (2008): why is AC a theorem of Tarski-Grothendieck set theory? It must have something to do with the universe axiom, but it's not clear what that axiom has to do with choice.
- Participants: A. Trybulec (long-time lead developer of the **mizar** project), J. Urban (formalizer of some relevant theorems and definitions), R. Solovay (critic).
- Result: a critique of the metamathematical organization of the part of the MML dealing with (strongly inaccessible) cardinals.
- Our fine-grained dependency graph can shed some light on this discussion and sharpen the critique of the MML.



# Influence of the universe axiom

- Trybulec's answer: look at **this formalized theorem of the MML**

theorem

for  $M$  being non countable  $\aleph$  holds

$M$  is strongly\_inaccessible implies Rank  $M$  is Tarski

- Rank  $M$  is **mizar**-ese for  $V_M$ , the  $M$ th stage of the von Neumann hierarchy.
  - a set  $X$  is Tarski when it is closed under powerset, taking subsets, and the condition that any subset of it that is not equipotent with it belongs to it.
- Tarski classes are well-orderable. Therefore, etc.



# Independence of universe axiom

- Claim: the formalized theorem doesn't depend on the universe axiom.
- Taken literally, this is quite incorrect. The universe axiom of **TG** is responsible for (among other things):
  - powerset
  - infinity
  - AC
- Anything in **TG** that uses either powerset, infinity, or choice depends on the universe axiom. Both powerset and infinity are used (directly or indirectly) by the formalized theorem.
- We can confirm that the formalized theorem by working with our dependency graph.



# Independence of the universe axiom: II

- However, by inspecting the dependency graph for the **mizar** Mathematical Library, we find that all paths of dependence from the formalized theorem to the universe axiom pass through either:
  - powerset,
  - infinity, or
  - choice
- Thus, although the **mizar** deduction  $d$  of the formalized theorem  $\varphi$  in question uses the universe axiom ( $d \mathbf{TG} \vdash \varphi$ ), we find that  $d$  contains a witness  $d^*$  to more interesting provability judgment:  $d^* \mathbf{ZFC} \vdash \varphi$ .
- This is a simple example of *(formal) proof refinement*.

# Dependence of the theorem in question on AC

Actually, Trybulec's suggested formalized theorem cannot explain why AC is a theorem of **TG**, because this proof depends on Zermelo's well-ordering principle! Here is a path of dependence:

1. **The theorem in question**
2. **Definition of von Neumann rank of a set**
3. **Every set is included in  $V_\alpha$ , for some ordinal  $\alpha$**
4. **Definition of a Tarski class**
5. **A sandwich theorem for cardinalities: if  $X \subseteq Y$  and  $Y \subseteq Z$ , and  $|X| = |Z|$ , then  $|X| = |Y|$  and  $|Y| = |Z|$**
6. **Definition of the cardinality operation on sets**
7. **Every set is equipotent with some ordinal**
8. **Zermelo's well-ordering principle**



# Advancing the critique: dependence of foundation

- Solovay showed how the formalized theorem depends on the axiom of foundation. He asserts that Tarski's proof of AC doesn't (essentially) depend on foundation.
  - We can confirm, with our dependency graph, that indeed the formalized theorem depends, indirectly, on foundation.
1. **Zermelo's well-ordering principle**
  2. **Redefinition of the subset relation on ordinals as membership**
  3. **Membership trichotomy theorem for ordinals**
  4. **Members of ordinals are themselves ordinals**
  5. **A trichotomy principle for membership:  $x \in y$  or  $y \in x$  or  $x \cap y = \emptyset$**
  6. **Axiom of foundation: every non-empty set  $x$  has a member that is disjoint from  $x$**

# Multiple ways of depending on foundation

- With our dependency graph, we find not just one path from the formalized theorem to foundation, but several.
- There are thus multiple *ways* in which foundation is needed.
  1. **Zermelo's well-ordering principle**
  2. **The inclusion relation restricted to an ordinal is a well-ordering**
  3. **The inclusion relation restricted to an ordinal is a connected relation**
  4. **Redefinition of the subset relation on ordinals as membership**
  5. **Membership trichotomy theorem for ordinals**
  6. **If a transitive set is a proper subset of an ordinal, then it is a member of the ordinal**
  7. **Axiom of foundation: every non-empty set  $x$  has a member that is disjoint from  $x$**

# Conclusion

- Working with a fine-grained dependency graph of the **mizar** Mathematical Library, one can find what is truly needed (minimally sufficient) for a given **mizar** theorem, definition, etc.
- We have seen one example of dependency graph-driven theory exploration (proof refinement, moving from **TG** to **ZFC**).

Future work:

- Working with automated theorem provers, we can augment the dependency information, proceeding further, potentially, to discovering *necessary* conditions for theorems, or possible *alternative proofs* that avoid what appears to be needed.



# References

- **The mizar homepage**
- **The mizar-items site.** (Explore the MML, including the items mentioned here, and find paths of dependence.)
- Solovay, Robert M., Re: AC and strongly inaccessible cardinals, posted to the Foundations of Mathematics mailing list.
- A. Quaife, *Automated Development of Fundamental Mathematical Theories*, Kluwer Academic Publishers, 1992
- S. Simpson, *Subsystems of Second-Order Arithmetic*, 2nd edition, Cambridge University Press, 2010
- J. Urban, “MPTP—Motivation, implementation, first experiments”, *Journal of Automated Reasoning*, **33**(3-4) (2004), pp. 319–339